Predicting Online Sales from Webpage Analytics

Lesley Miller

24/01/2020

Contents

Description of the variables, data source

Table 1: Distribution of Special Day

S No.	Variable	Description
1	`Administrative`	Number of Distinct administrative pages
2	`Informational`	Number of Distinct Informational pages
3	`ProductRelated`	Number of Distinct Product Related pages
4	`Administrative_Duration`	Time(in seconds) spent on Administrative pages
5	`Informational_Duration`	Time(in seconds) spent on Informational pages
6	`ProductRelated_Duration`	Time(in seconds) spent on Product Related pages
7	`BounceRates`	Average bounce rate of all web-pages visited by user. For a web-page its the percentage of people who visit the website from that webpage and left without raising any other request

S No.	Variable	Description
8	`ExitRates`	Average exit rate of all web-pages visited by user: For a web-page its the percentage of people who exited the website from that webpage
9	`PageValues`	Average page value of all web-pages visited by user: For a web-page its the average dollar-value of that page which the user visited before completing the transaction
10	`SpecialDay`	The closeness of site visitng time to a special day (higher chances of a session resulting in a transaction)
11	`OperatingSystems`	Operating system used by the user
12	`Month`	Month of Year
13	`Browser`	Browser used by the user
14	`Region`	Geographic region
15	`TrafficType`	Type of Channel user by the user to arrive at the website
16	`VisitorType`	Type of the visitor
17	`Weekend`	Weekend indicator
18	`Revenue`	Revenue transaction indicator

Table 2: Distribution of Month

Special Day	Percentage of Shopping Sessions
0	89.8
0.2	1.3
0.4	2.0
0.6	3.0
0.8	2.5
1	1.3

Table 3: Distribution of Operating Systems

Month	Percentage of Shopping Sessions
Aug	3.3
Dec	14.0
Feb	1.4
Jul	3.5
June	2.5
Mar	15.4
May	27.5
Nov	24.5
Oct	4.5
Sep	3.5

Table 4: Distribution of Browser

Operating Systems	Percentage of Shopping Sessions
Aug	3.3
Dec	14.0
Feb	1.4
Jul	3.5
June	2.5
Mar	15.4
May	27.5
Nov	24.5
Oct	4.5
Sep	3.5

Table 5: Distribution of Region

Browser	Percentage of Shopping Sessions
1	20.0
2	64.5
3	0.8
4	5.8
5	3.8
6	1.4
7	0.4
8	1.2
9	0.0
10	1.4
11	0.1
12	0.1
13	0.5

Table 6: Distribution of Traffic Type

Region	Percentage of Shopping Sessions
1	38.8
2	9.0
3	19.4
4	9.6
5	2.7
6	6.7
7	6.3
8	3.3
9	4.1

Table 7: Distribution of Visitor Type

Traffict Type	Percentage of Shopping Sessions
1	19.7
2	32.1

Traffict Type	Percentage of Shopping Sessions
3	16.4
4	8.7
5	2.1
6	3.5
7	0.3
8	2.9
9	0.3
10	3.7
11	2.0
12	0.0
13	6.0
14	0.1
15	0.3
16	0.0
17	0.0
18	0.1
19	0.2
20	1.5

Table 8: Distribution of Weekend

Visitor Type	Percentage of Shopping Sessions
New_Visitor	13.9
Other	0.7
Returning_Visitor	85.4

Table 9: Distribution of Weekend and Revenue

Weekend	Percentage of Shopping Sessions
FALSE	76.4
TRUE	23.6

Table 10: Cumulative Distribution of Numberic Variables

Revenue	Percentage of Shopping Sessions
FALSE	84.1
TRUE	15.9

Table 11: Cumulative Distribution of Numberic Variables

probs	Administrative_Duration	Informational_Duration	ProductRelated_Duration
0.00	0.000	0.000	0.000
0.05	0.000	0.000	0.000
0.10	0.000	0.000	36.000
0.15	0.000	0.000	77.000
0.20	0.000	0.000	128.000

probs	Administrative_Duration	Informational_Duration	ProductRelated_Duration
0.25	0.000	0.000	181.500
0.30	0.000	0.000	243.983
0.35	0.000	0.000	315.000
0.40	0.000	0.000	402.009
0.45	0.000	0.000	493.640
0.50	8.000	0.000	598.600
0.55	22.971	0.000	718.855
0.60	38.000	0.000	855.133
0.65	54.500	0.000	1029.246
0.70	73.020	0.000	1229.497
0.75	95.763	0.000	1473.834
0.80	124.640	0.000	1784.082
0.85	165.388	25.185	2210.075
0.90	228.320	70.000	2905.811
0.95	354.752	193.463	4322.162
1.00	3398.750	2549.375	63973.522

probs	BounceRates	ExitRates	PageValues	ProductRelated
0.00	0.000	0.000	0.000	0
0.05	0.000	0.005	0.000	1
0.10	0.000	0.007	0.000	3
0.15	0.000	0.010	0.000	4
0.20	0.000	0.012	0.000	6
0.25	0.000	0.014	0.000	7
0.30	0.000	0.016	0.000	9
0.35	0.000	0.018	0.000	11
0.40	0.000	0.020	0.000	13
0.45	0.000	0.023	0.000	15
0.50	0.003	0.025	0.000	18
0.55	0.005	0.029	0.000	21
0.60	0.007	0.032	0.000	24
0.65	0.010	0.036	0.000	28
0.70	0.013	0.041	0.000	32
0.75	0.017	0.050	0.000	38
0.80	0.023	0.058	3.454	45
0.85	0.033	0.073	9.927	57
0.90	0.059	0.100	19.227	74
0.95	0.200	0.200	38.880	111
1.00	0.200	0.200	361.764	705

Project Summary

Why it Matters

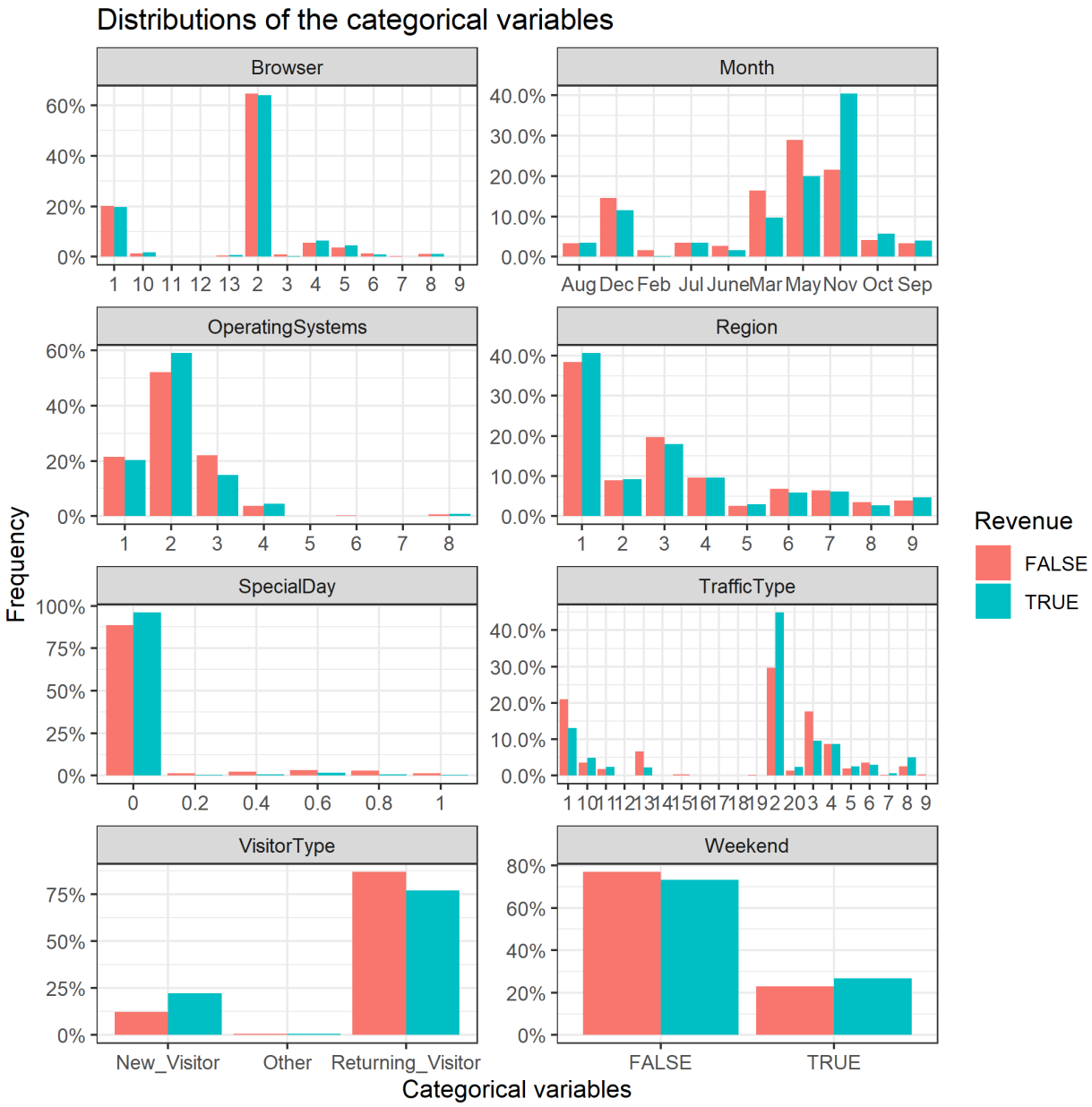
Methods

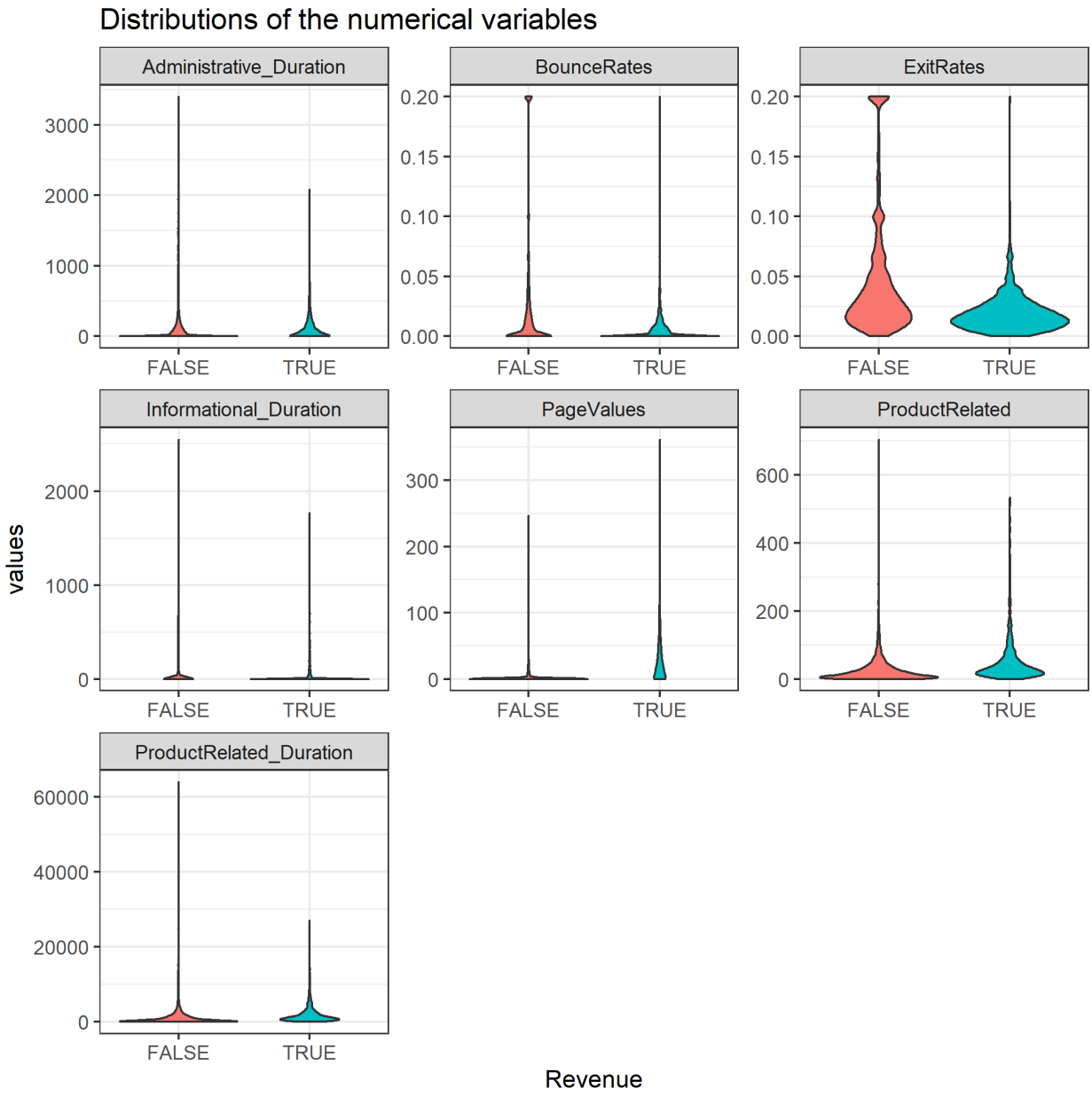
Data Source

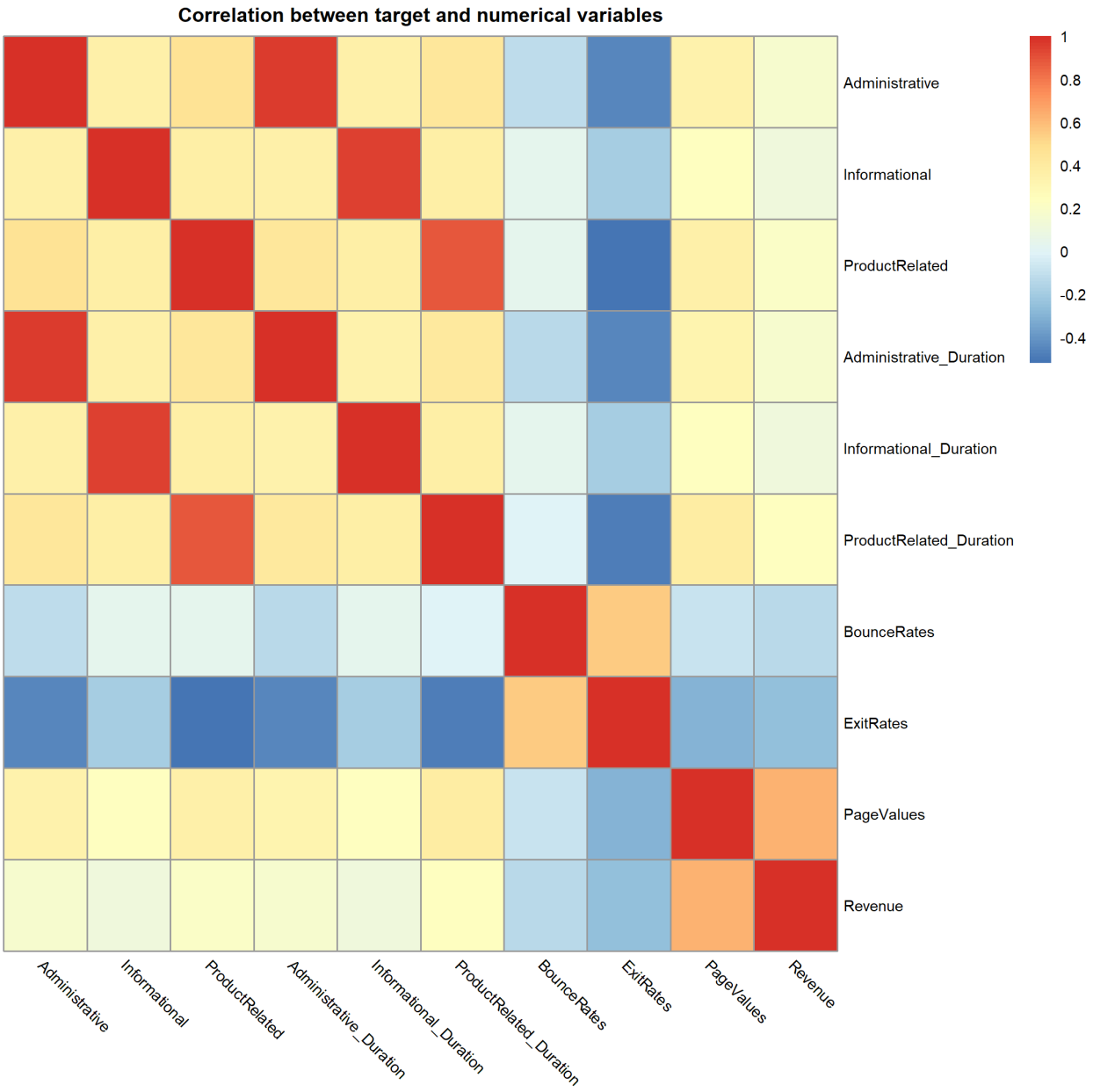
turn this into written text not bullent points.

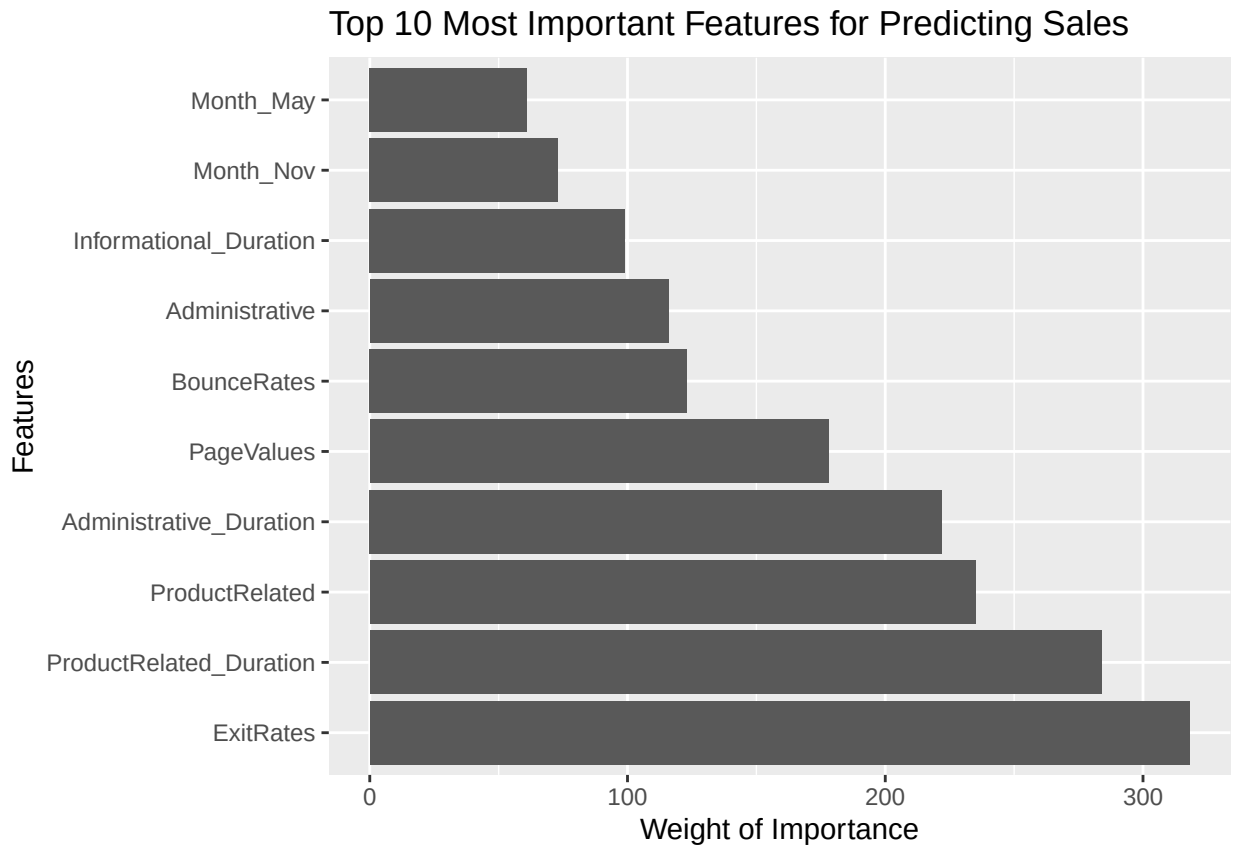
- Each row represent a session by a user.
- Each user has only 1 session in the dataset.
- The data is for 1-year period.
- ~15% sessions resulted in a purchase.

Exploratory Data Analysis




Distributions of the numerical variables
Administrative_Duration
BounceRates
ExitRates
Informational_Duration
PageValues
ProductRelated
ProductRelated_Duration
values
Revenue
FALSE
TRUE


Correlation between target and numerical variables
Administrative
Informational
ProductRelated
Administrative_Duration
Informational_Duration
ProductRelated_Duration
BounceRates
ExitRates
PageValues
Revenue
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4



Modelling

Results and Discussion

Limitations

Future Directions

References